

# Intersubtype Recombination in HIV-1 and HIV-2

---

Beatrice H. Hahn<sup>1</sup>, David L. Robertson<sup>2</sup>, and Paul M. Sharp<sup>2</sup>

<sup>1</sup> *Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA*

<sup>2</sup> *Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK*

## Introduction

Recombination has long been known to represent an important means by which retroviruses generate genetic diversity. Each retrovirus particle contains a dimeric RNA genome and a reverse transcriptase enzyme that can switch templates during proviral DNA synthesis. Several mechanisms have been proposed to explain the recombination process, all of which are believed to operate during the reverse transcription step [1]. Exchange of genetic material between divergent viruses is thus only possible if their genomes replicate in the same cell and are packaged into the same virion. Although recombinational events are frequent under laboratory conditions, recombination has not been thought to represent a significant source of new variation for the Human Immunodeficiency Viruses. Until very recently, individuals have not been found to be coinfecting with multiple divergent strains of HIV, and recombination has been assumed to only involve the rather closely related members of the viral quasispecies that evolves over the course of infection. However, as more and more isolates are molecularly characterized, evidence is increasing for the existence of mosaic HIV-1 and HIV-2 genomes which appear to have resulted from recombination events involving rather divergent viruses.

## Mosaic HIV Proviruses

Globally circulating HIV-1 and HIV-2 strains exhibit extensive genetic diversity (see other chapters in this compendium). Phylogenetic analyses have revealed two distinct “groups” (M and O) for HIV-1, as well as (at least) nine “sequence subtypes” (A-I) within the major group M. Five sequence subtypes (A-E) have also been identified for HIV-2, although these are based on the analysis of considerably fewer viruses [2]. Importantly, the majority of HIV-1 and HIV-2 viruses cluster consistently in phylogenetic trees derived from different regions of their genome. Nevertheless, recent studies have noted several viruses which occupy different positions in phylogenetic trees, depending on which gene or part of the genome is used for analysis [3-9]. For example, one of the earliest characterized HIV-1 isolates from Africa (MAL) was found to cluster with subtype D viruses in *env*, but to group with subtype A viruses in *gag* [3,4]. Similarly, HIV-1 isolates from Thailand were found to form a distinct sequence subtype in *env* (E), but to group with subtype A viruses in *gag* [5,6]. Finally, one strain of HIV-2 (7312A) was identified to cluster with subtype A viruses in *env*, but to group with subtype B viruses in *gag* [2,7]. These findings suggested that individuals can become coinfecting with highly divergent strains of HIV, and that recombination between their genomes can generate biologically active viruses.

To further examine these potential recombinants, we have recently characterized the evolutionary histories of the HIV-1MAL and HIV-27312A genomes in greater detail [4]. Importantly, both proviruses represent viral genomes cloned as a single genetic unit by lambda phage cloning techniques. Moreover, both proviruses have been sequenced in their entirety. As a first step, we constructed evolutionary trees from complete gene (nucleotide or protein) sequences. Since recombination events are not expected to be localized to intergenic regions, we next estimated phylogenies for partial gene sequences, looking for discordant branching orders among trees from different parts of the gene. Finally, to map more precisely the putative recombination cross-over points, we inspected the linear distribution of phylogenetically informative sites supporting alternative tree topologies. This was done by considering an alignment of just four sequences, consisting of the putative recombinant, one (non-recombinant) sequence (preferably a consensus sequence) representing each of the two subtypes seemingly involved in the recombination event, and an outgroup. Breakpoints were inserted at each possible point between adjacent informative sites and a  $2 \times 2$  heterogeneity  $\chi^2$  was calculated for the numbers of sites supporting the clustering of

the putative recombinant with each of the consensus sequences. The likely breakpoint was identified as that which gave the maximal  $\chi^2$  value. To assess the probability of obtaining by chance as high a  $\chi^2$  value as that observed, 10,000 simulations were performed in which the informative sites were randomly shuffled (for more details see refs. 4 and 10).

The results of these analysis revealed that both HIV-1MAL and HIV-27312A contained recombinant genomes of considerable genetic complexity. Both proviruses exhibited evidence for multiple recombination cross-overs. There were at least five breakpoints in HIV-1MAL delineating six genomic regions with different phylogenetic histories (Figure 1). Three of these (the 5' part of gag, the 3' part of pol plus the 5' half of vif, and the extreme 3' part of env plus nef) were subtype A-like (for example, clustering with U455). Two others (the 3' part of gag and the 3' half of vif through env) were subtype D-like (for example, clustering with NDK and ELI). The remaining region (5' part of pol) was not obviously related to either subtype A or D sequences (clustering outside U455, NDK and ELI). Analysis of HIV-27312A yielded results of a similar nature (Figure 2). There were four different points of cross-over, but in this case all of these mapped to the env gene. Most of the 5' part of the 7312A genome (gag through the first exon of tat/rev) was subtype B-like (clustering with D205 and UC1). By contrast, most of the env gene was subtype A-like (clustering with ROD and ST). Finally, there was a 300 - 400 bp region within gp120 that was subtype B-like. The mosaic nature of 7312A was also documented in vivo, since env and gag sequences derived by PCR from primary patient material and sequential blood samples yielded the same branching order [2,7].

### HIV-1 Intersubtype Recombinants in the Database

To examine whether the mosaic nature of HIV-1MAL and HIV-27312A was representative of other HIV infections, we screened the HIV-1 sequence database for additional examples of intersubtype recombination [10,11]. All HIV-1 viruses for which near full-length gag and env sequences were published or deposited in GenBank, were selected for this analysis. Again, phylogenetic trees were constructed using various regions of these gag and env sequences. Putative recombinants were identified as those viruses whose sequences fell in significantly discordant positions in different trees, and these sequences were then subjected to further breakpoint analysis. The results are shown in Figure 3. Among 150 viruses analyzed, 17 (including MAL) were found to be mosaic, containing sequences from eight different group M subtypes (Figure 3 also includes a recently published Brazilian B/F strain for which only gp120 sequences are available; ref. 8). As previously seen with HIV-1MAL and HIV-27312A, most gag and env genes contained multiple points of cross-over, suggesting that viruses with hybrid gag and env proteins are viable. In addition, there were a number of regions like the 5' part of pol in HIV-1MAL for which subtype classification was not possible (denoted by question marks in Figure 3). Such regions may be highly mosaic with several internal cross-overs which are too complex to be identified. Alternatively, they may contain sequences from new subtypes which are not yet represented in the database. Importantly, there was no evidence for recombination between group M and group O viruses.

Figure 3 contains only one representative of viruses from Thailand which have been classified as belonging to "subtype E" (CM240). Members of this subtype are all closely related, presumably reflecting a founder effect. Moreover, members of this subtype have long been known to branch inconsistently in phylogenetic trees derived from different parts of their genomes. These viruses form a distinct phylogenetic cluster (termed subtype E) in env trees, but their gag sequences fall into subtype A [5,6]. In addition, recent examination of full length gp160 sequences of these viruses has indicated that their env genes are mosaic [11]. Sequences from the 3' half of the "subtype E" gp41 coding domain also cluster in subtype A. These findings are consistent with the hypothesis that all "subtype E" viruses are in fact A/E intersubtype recombinants. However, recombination breakpoint analysis is not possible because, as of yet, there is no sequence available for any virus known to be a full length representative of subtype E. Thus, definitive analysis must await additional sequence information.

It is also important to note that all of four "subtype G" viruses currently deposited in the database appear to be mosaic (Figure 3). GA-VI525 is recombinant (as a whole) because its env sequences clusters in subtype G, whereas its gag sequence clusters in subtype H [10,12]. The remaining three "subtype G" viruses UG975, RU131, and LBV21-7 have mosaic env genes [11]. Although most of their env coding region falls into subtype G, sequences in the 3' half of their gp41 domain cluster in subtype

## Recombination in HIV-1 and HIV-2

A. Again, we cannot define the recombination cross-over points with certainty, primarily because of lack of sufficient subtype G sequences in the database. However, our preliminary data would suggest that they are near the membrane spanning domain, and possibly at the same location in all three viruses.

Finally, other investigators have also identified examples of intersubtype recombination in the database using somewhat different approaches and methodologies [13,14]. Siepel and colleagues have recently screened partial env (total of 211) and gag (total of 85) sequences with a newly developed Recombinant Identification Program (RIP, ref. 13). Among newly released database entries, they have identified five additional intersubtype recombinants, including a mosaic gp120 sequence comprised of segments belonging to subtypes A and E (CAR4039). Salminen and coworkers analyzed database gag and env sequences with their "Bootscreening" Method and have suggested that there is an additional segment of subtype F origin within the 5' region of the BZ200 gag gene [14]. All three approaches that have been used to find recombinant viruses (i.e., our methodology, and those of Siepel et al. and Salminen et al.) look for differences in phylogenetic signal along an alignment. All three have produced very similar results, although comparisons suggest that our approach may be somewhat more conservative in identifying mosaic regions.

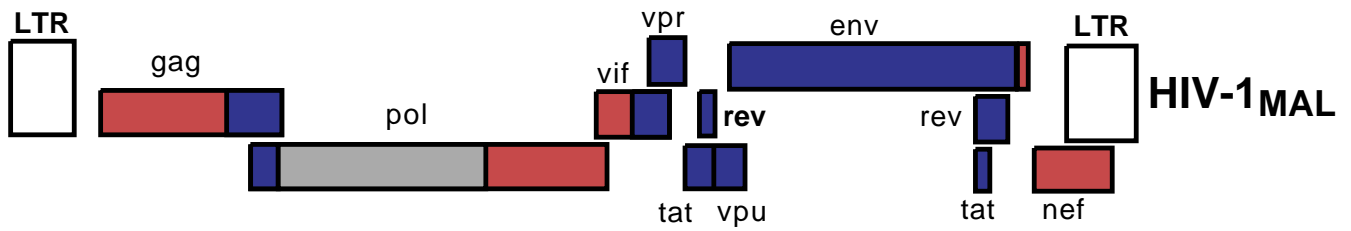
## Recombination in SIV

It should also be emphasized that recombination involving divergent viruses is not limited to primate lentiviruses infecting humans. Analysis of a complete proviral sequence of an SIVAGM from a sabaeus monkey revealed discordant phylogenetic relationships from analyses of different regions of its genome [15]. A tree based on the complete env protein sequence confirmed that SIVAGMsab represents a distinct lineage within the SIVAGM radiation, approximately equidistant from SIVAGM infecting vervet (SIVAGMver) and grivet (SIVAGMgri) monkeys. In contrast, in trees from the 3' gag and the 5' pol region SIVAGMsab clustered with the HIV-2/SIVSM lineage (Figure 4). The 5' half of gag did not appear to be more closely related to either SIVAGM or SIVSM, no matter how the region was subdivided for analysis. A second independent sabaeus isolate exhibited similar phylogenetic relationships (not shown). These results indicate that sabaeus monkey viruses comprise a mosaic genome which must have resulted from recombination of divergent lentiviruses in the past.

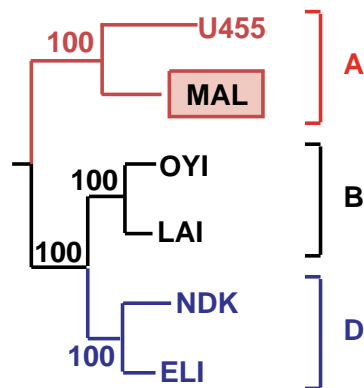
## Implications of Recombination

The examples of HIV-1MAL and HIV-27312A, along with the surprising number of mosaic HIV-1 sequences in the database, imply that a significant proportion of individuals have become co-infected (or super-infected) with HIV strains belonging to different sequence subtypes. This finding has immediate consequences for HIV-1 vaccine development. First of all, it raises questions concerning the timepoints and circumstances under which individuals acquire divergent strains of HIV. Coinfection has been thought to be rare and consequently it has been assumed that the host's immune response plays a role in protecting against superinfecting viruses. This may not be so, particularly when the superinfecting strain is highly divergent, i.e. belongs to a different sequence subtype. Alternatively, the mosaic genomes identified in this study could be the result of simultaneous exposure to different viruses or superinfection during the first weeks or months, before an effective immune response is mounted. Clarification of these issues is needed to gain insights into correlates of immune protection or failure.

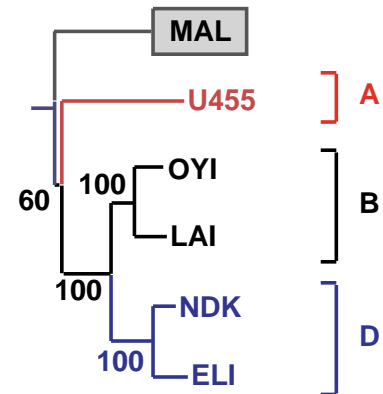
In the context of ongoing vaccine development efforts, in which vaccine preparations are aimed at individual sequence subtypes (e.g., planned efficacy trials in Thailand), it is also clear that diversity surveys must include measures to identify mosaic viruses. Indeed, it can be assumed that recombinants are generated in all areas where divergent subtypes spread simultaneously through the same population. Future studies will need to address whether genetic changes of the type reported above can influence clinically important properties of HIV, such as transmissibility, virulence, and replication potential. Finally, it is important to realize that our present analysis almost certainly underestimates the frequency of coinfections, since recombinants of divergent viruses from the same sequence subtype are much more difficult to detect.



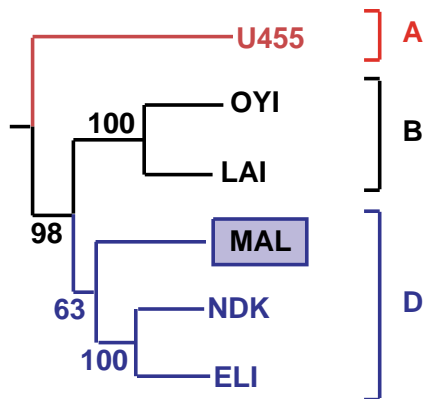
5' gag



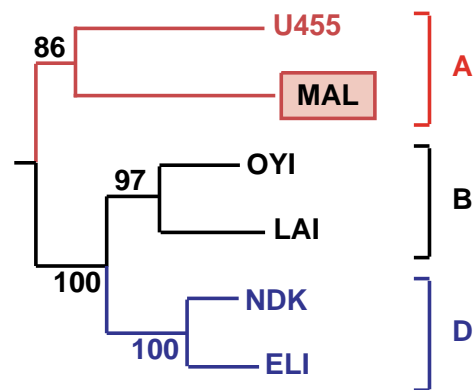
5' pol



env



nef



10% divergence

Figure 1. Schematic representation of the mosaic genome of HIV-1MAL (adapted from ref. 4).

Recombination in HIV-1 and HIV-2

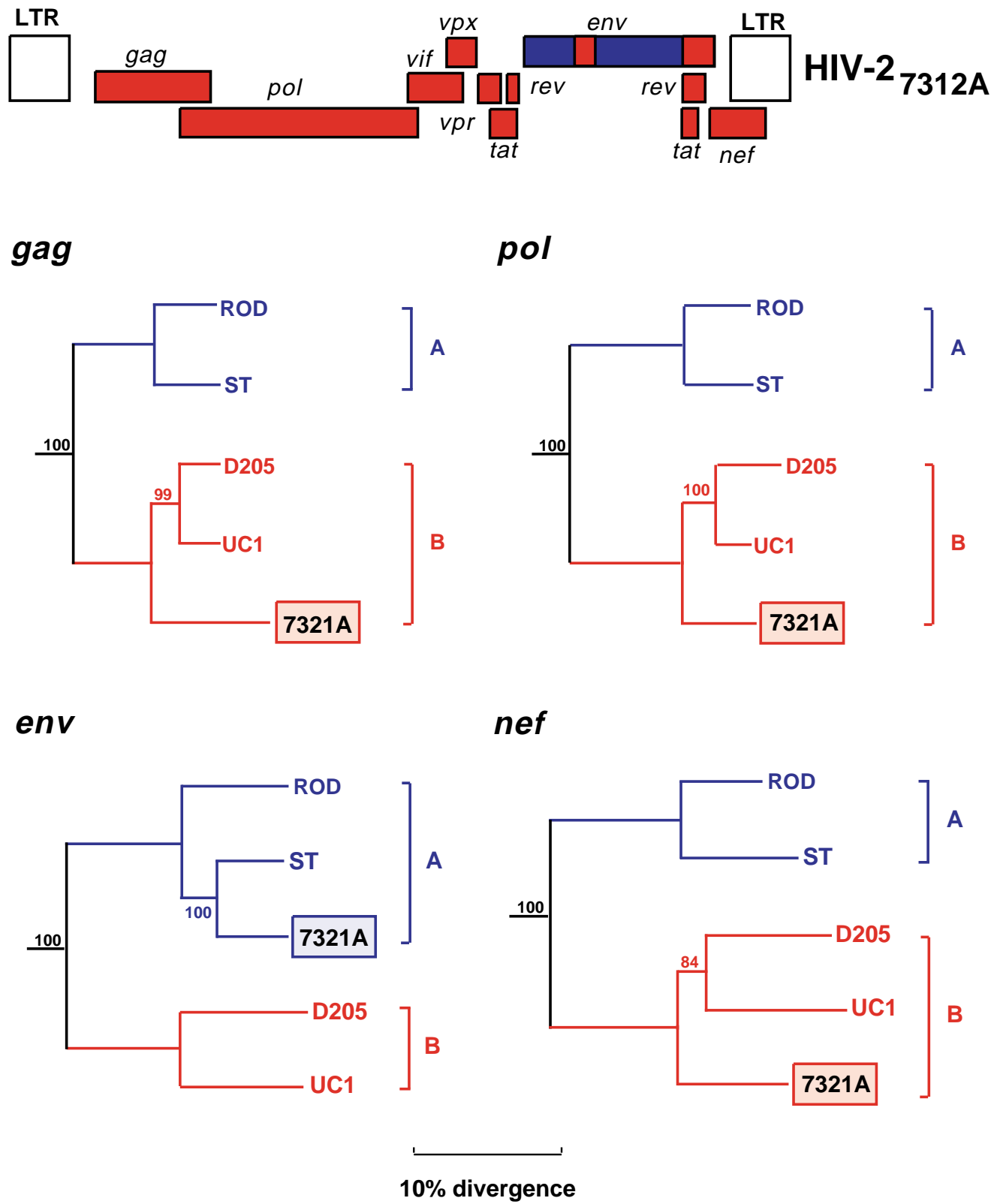


Figure 2. Schematic representation of the mosaic genome of HIV-27312A (adapted from ref. 4).

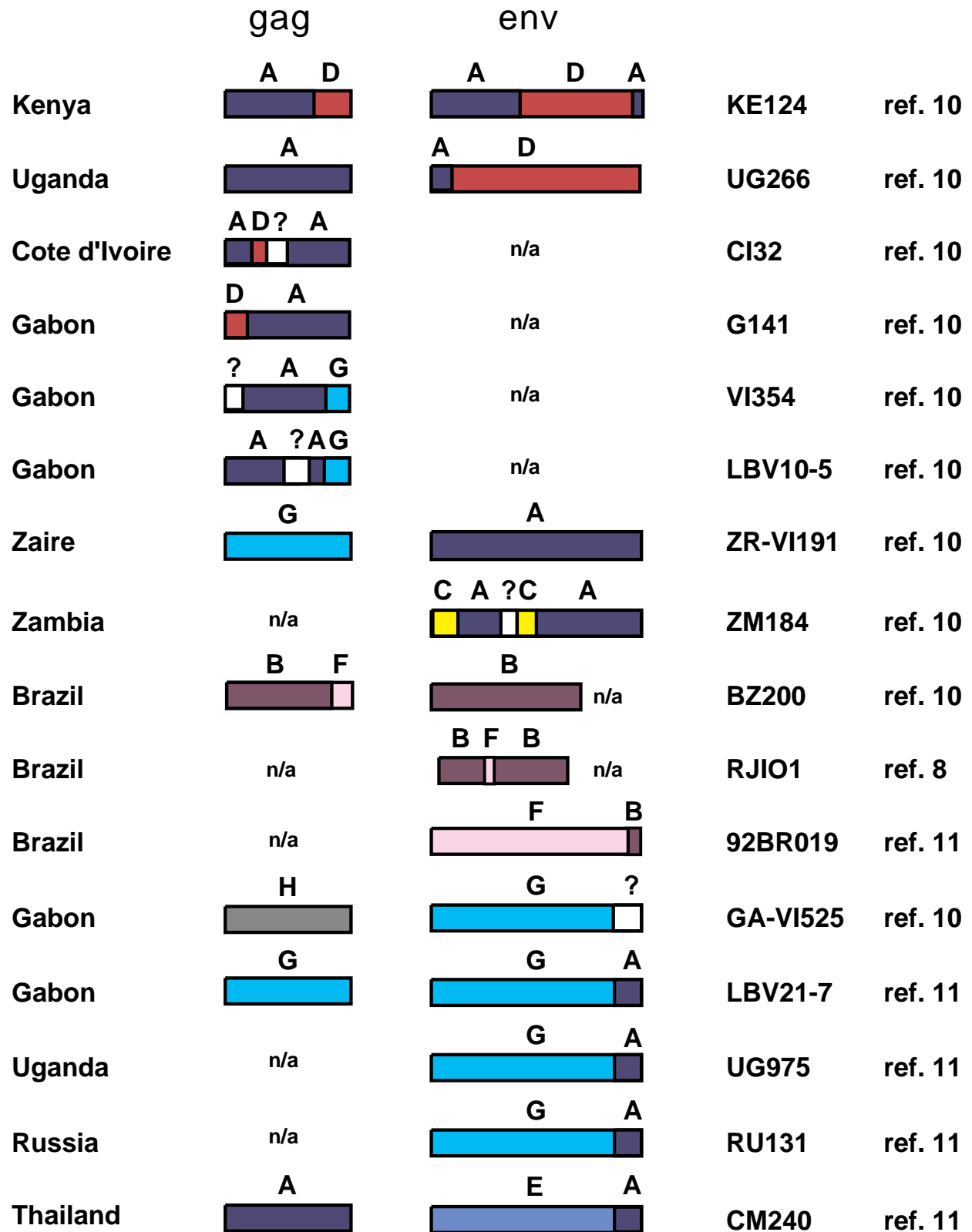


Figure 3. Recombinant HIV-1 gag and env sequences in the database (adapted from refs. 10 and 11).

Recombination in HIV-1 and HIV-2

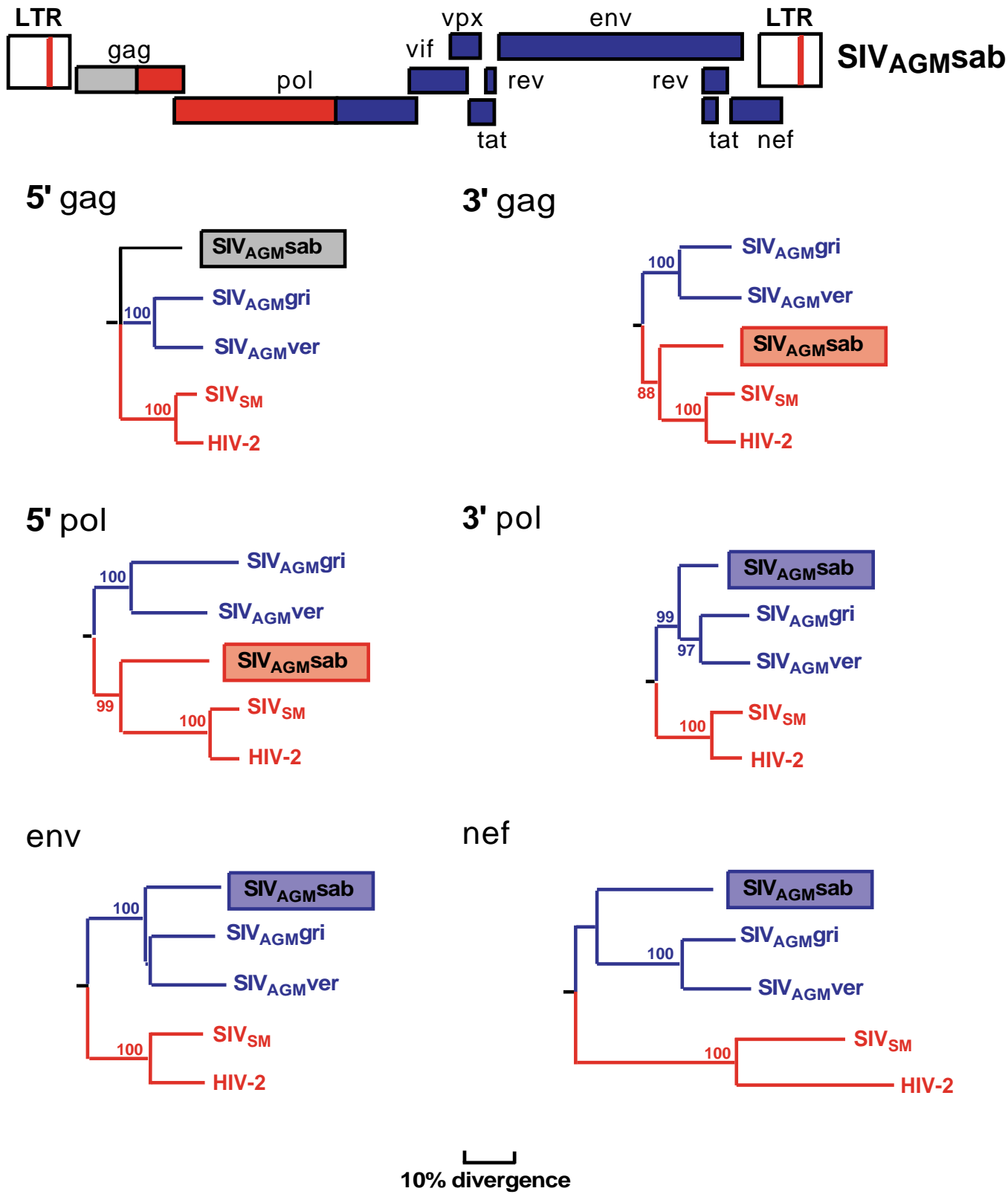


Figure 4. Schematic representation of the mosaic genome of SIVAGMsab (adapted from ref. 15)

## References

- [1] Coffin, J. M. Retroviridae and their replication. In: B. N. Fields and D. M. Knipe (eds), *Fields Virology*. Raven Press, p. 1437–1500, 1990.
- [2] Gao F., Yue L., Robertson D.L., Hill S.C., Hui H., Biggar R.J., Neequaye A.E., Whelan T.M., Ho D.D., Shaw G.M., Sharp P.M., and Hahn, B.H. Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. *J. Virol.* **68**: 7433–47, 1994.
- [3] Li W.-H., Tanimura M., and Sharp P.M. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**:313–330, 1988.
- [4] Robertson, D. L., B. H. Hahn, and P. M. Sharp. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249–259, 1995.
- [5] McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, and D. S. Burke. Genetic variants of HIV-1 in Thailand. *AIDS Res. Hum. Retroviruses* **8**:1887–1895, 1992.
- [6] Sharp P.M., Robertson D.L., Gao F., and Hahn B.H. Origins and diversity of Human Immunodeficiency Viruses. *AIDS* **8**: S27–42, 1994.
- [7] Gao, F., L. Yue, A. T. White, P. G. Pappas, J. Barchue, A. P. Hanson, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn. Human infection by genetically diverse SIV related HIV-2 in West Africa. *Nature* **358**:495–499, 1992.
- [8] Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. M. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Gálvao-Castro, J. I. Mullins, and A. Mayer. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J. Virol.* **68**:6340–6346, 1994.
- [9] Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brostrom, H.B. Hansson, M. Uhlen, and J. Albert. Biological and molecular characterization of subtype D,G, and A/D recombinant HIV-1 transmissions in Sweden. *Virology* **209**:136–146, 1995.
- [10] Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn. Recombination in HIV-1. *Nature* **374**:124–126, 1995.
- [11] Gao, F., S. G. Morrison, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, D. L. Robertson, P. M. Sharp, G. M. Shaw, B. H. Hahn, and WHO and NIAID Networks for HIV Isolation and Characterization. Molecular cloning and analysis of functional envelope genes from HIV-1 sequence subtypes A through G, submitted.
- [12] Janssens, W., L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J.-L. Perret, C. Atende, P. Piot, and G. Van Der Groen. Genetic and phylogenetic analysis of env subtypes G and H in central Africa. *AIDS Res. Hum. Retroviruses* **10**:877–879, 1994.
- [13] Siepel, A.C., A.L. Halpern, C. Macken, and B.T.M. Korber. A computer program designed to rapidly screen for HIV-1 intersubtype recombinant sequences. *AIDS Res. Hum. Retroviruses* **11**:1413–1425, 1995.
- [14] Salminen, M.O., J.K. Carr, D.S. Burke, and F.E. McCutchan. Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11**:1423–1425, 1995.
- [15] Jin, M. J., H. Hui, D. L. Robertson, M. C. Muller, F. Barre-Sinoussi, V. M. Hirsch, J. S. Allan, G. M. Shaw, P. M. Sharp, and B. H. Hahn. Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. *EMBO Journal* **13**:2935–2947, 1994.